

## Content-as-a-service platform with the alfresco open-source enterprise content management system

### Siang Hock Kia

Technology & Innovation, National Library Board, Singapore  
siang\_hock\_kia@nlb.gov.sg

### Zhi Liang Wang

Technology & Innovation, National Library Board, Singapore  
zhi\_liang\_wang@nlb.gov.sg



Copyright © 2013 by **Siang Hock Kia & Zhi Liang Wang**. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License:

<http://creativecommons.org/licenses/by/3.0/>

---

### Abstract:

*The National Library Board (NLB) of Singapore embarked on a holistic programme over the last few years to transform the access and discovery of digital library resources for the 'connected generation'. Uniquely Singapore materials are collected, both from institutions and man-on-the-street, and if necessary digitized for easy access.*

*A flexible, scalable, robust, and yet cost-effective content management service (CMS) is critical to the success of the programme. To enable maximum reach for the rich-media content, the CMS has been envisaged from the inception as a 'content-as-a-service' platform that is loosely-coupled with the online services that would consume the content via application programming interfaces (APIs). This service architecture has allowed NLB to develop content micro-sites that are search engine friendly and optimized for multi-screens on all devices with the responsive web design framework.*

*NLB selected the Alfresco Community Edition, an open source Enterprise Content Management system, as its content management platform in 2010. The platform has been successfully used in key content-rich NLB services, including Infopedia, BookSG, PictureSG, MusicSG, HistorySG, the Standards Portal and the Singapore Memory Portal.*

**Keywords:** Open Source Software, Content-as-a-service, Alfresco Enterprise Content Management System, Application Architecture, National Library Board (NLB) of Singapore.

---

## **1 INTRODUCTION**

Singapore is one of the most connected nations in the world with household broadband and mobile penetration at 105% and 152% respectively as at December 2012. Singaporeans, especially the young, are increasingly moving into the digital spaces to meet their social, information and learning needs.

Recognising these megatrends, the National Library Board (NLB) of Singapore embarked on a holistic programme over the last few years to transform the access and discovery of digital library resources for the 'connected generation'. The three pillars of the programme are content, findability and social+mobility.

### **Content**

Overseeing the National Library in Singapore, NLB is in the strategic position to effect the transformation. Where 'content is king' in the Internet era, NLB manages rich and comprehensive collections of information that are of historical and cultural significance to Singapore.

It can also better engage and synergise with the private and public sectors. A good example off-cited is the NewspaperSG<sup>1</sup> service where NLB was able to reach win-win agreements with newspaper publishers in Singapore to allow significant amount of copyrighted newspaper articles to be made available online.

### **Findability**

With the popular search engines becoming the default starting point for people to look for information, it is important that the content which NLB avails online through its sustained acquisition and digitisation efforts can be found via these search engines.

The content-rich NLB services were designed from ground-up to be search engine friendly. This has resulted in a quantum increase in the access of the NLB content generated directly from searches via the popular search engines. For an example, over 60% of the traffic to the Infopedia<sup>2</sup> service (the electronic encyclopedia on Singapore's history, culture, people and events) comes directly from search engines. As a result, the total page views skyrocketed from just 400 to over 200,000 a month!

### **Social+Mobility**

Over 60% of the Singapore population are Facebook members. With the high mobile and social media penetration rates in Singapore, NLB places great emphasis in reaching our users through their social and mobile spaces.

---

<sup>1</sup> NewspaperSG is an online resource of current and historic Singapore and Malaya newspapers (<http://newspapers.nl.sg>)

<sup>2</sup> Singapore Infopedia is an online encyclopedia on Singapore's history, culture, people and event (<http://infopedia.nl.sg>)

The Library in Your Pocket<sup>3</sup>, Mobile Read<sup>4</sup> and the MyLibrary@Facebook<sup>5</sup> services enable our users to enjoy the comprehensive and rich library services and resources on-the-go, and even via their Facebook accounts.

## 2 CONTENT-AS-A-SERVICE

Content is the foundation of the programme, as without relevant content, nothing matters. Uniquely Singapore materials are collected, both from institutions and men-on-the-street, and if necessary digitized for easy access.

A flexible, scalable, robust, and yet cost-effective content management service (CMS) is critical to the success of the programme. To enable maximum reach and flexibility for the delivery of the rich-media content, the CMS has been envisaged from the inception as a ‘content-as-a-service’ platform that is loosely coupled from the online services that would consume the content via application programming interfaces (APIs).

A clear demonstration of the agility achieved by NLB with this architectural decision has been the adoption of the Responsive Web Design (RWD) framework<sup>6</sup> recently. The RWD framework leverages on advances in the HTML5 and CSS3 standards and their supports in modern browsers on all the common platforms to enable websites to be designed to adapt their layout according to the form factor of the devices accessing them. Figure 1 illustrates the application of the framework to the Singapore Infopedia service.



Figure 1: The Singapore Infopedia service adapting to the form factors of the devices (Apple iPad, Samsung Galaxy Tablet, Sony Xperia V)

The RWD framework is a cost-effective approach to tackle the issue of the proliferation of the myriad of devices with which our library users are accessing the NLB resources. Typically, content management systems stipulate their own templating mechanism for user

<sup>3</sup> Library in your pocket (LiYP) is the mobile-friendly digital library of NLB (<http://m.nlb.gov.sg>)

<sup>4</sup> Mobile Read: Apple App Store - <http://itunes.apple.com/sg/app/mobileread/id411883163?mt=8>; Google Play Store: <https://play.google.com/store/apps/details?id=sg.gov.nlb.mobileread>

<sup>5</sup> MyLibrary @ Facebook: <https://apps.facebook.com/nlbmylibrary/>

<sup>6</sup> Responsive Web Design by Ethan Marcotte, 25 May 2010 (<http://alistapart.com/article/responsive-web-design>)

interface design and development. The loose-coupling of the online service and the content management service removes any constraint imposed by such templating system and provided NLB the flexibility to quickly adopt the RWD framework. Moreover, as NLB is not tied to the templating system, we can easily change the CMS without impacting the online services.

Section 6 describes other benefits achieved with the content-as-a-service approach.

### 3 KEY REQUIREMENTS OF THE CMS

Over the years, NLB has been progressively evolving a suite of application service components (collectively termed the Service Enablement Layer) to enable the delivery of innovative services to the library users. These re-usable components cover rich media content management, search & discovery, multi-channel content delivery, and personalisation & engagement (see Figure 2).

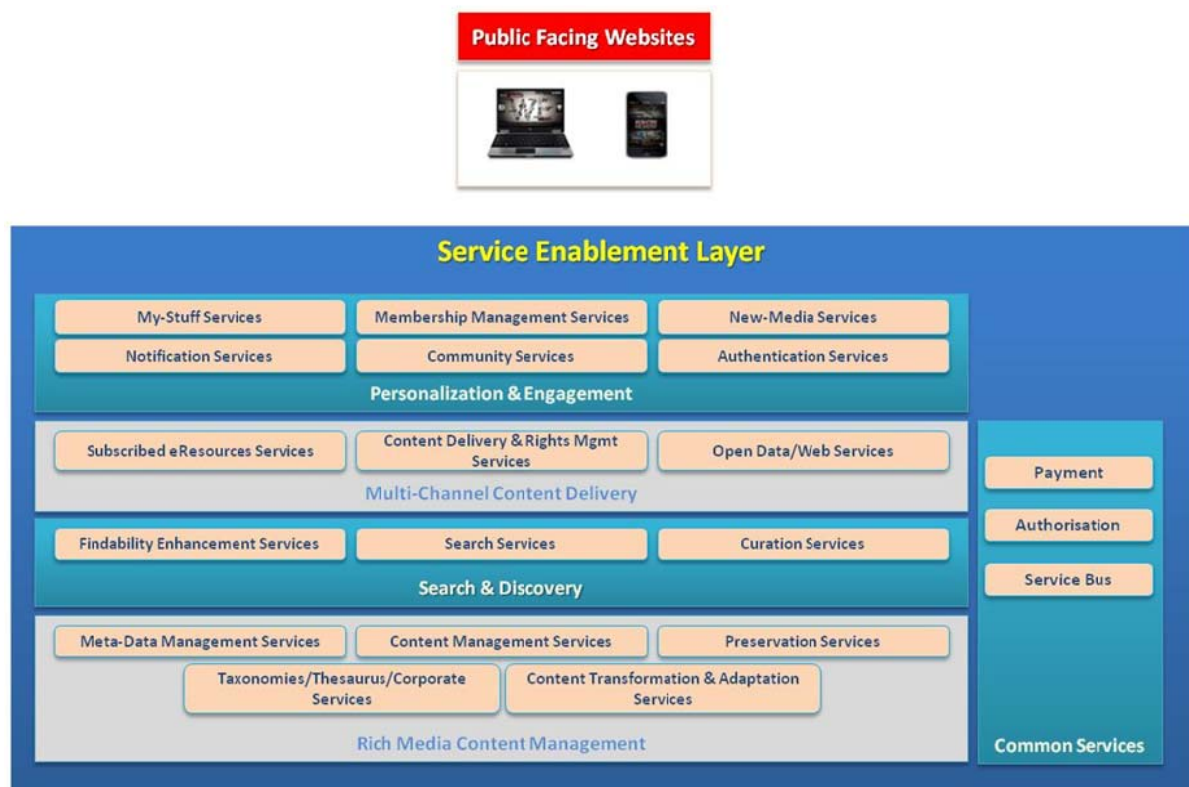


Figure 2: The NLB Service Enablement Layer

As a key foundational component of the Service Enablement Layer, the Content Management Service (CMS) has the following key requirements:

- Flexible data model. The CMS is a central repository that will manage the vast and myriad collections of the access copy of the NLB content, including HTML text, images, PDF, ePUB, music tracks, oral history, video, and many more. It is therefore important for the CMS to be built upon a flexible and extensible data model.

Changing the data model should not require source code and database changes, and source code re-compilation.

NLB has adopted the NLB Application Profile (NLBAP), a Dublin Core Application Profile (DCAP) as its internal standard to describe digital objects. The CMS needs to handle the NLBAP, and its future extensions.

- Extensive APIs. As the CMS is a content-as-a-service platform that is loosely coupled with the anticipated online services, openness is crucial. These online services will have full flexibility in the design of the user experience, while consuming the APIs provided by the CMS. The out-of-the-box APIs should be already extensive.
- High performance bulk ingest. Born-digital and digitised content typically get into the CMS on a batch basis as a result of upstream acquisition and/or digitisation activities. It is therefore important for the CMS to come with a high performance bulk ingest tool that would be able to load hundreds of thousands of meta-data and digital objects efficiently.
- Robustness. NLB users have come to expect excellent, responsive and highly available services. It is therefore critical that the CMS platform has been designed and built to operate in a demanding environment. Load balancing and high availability options should be available.
- Supportability. It is important that the CMS platform is well supported, either by the community or the developer, preferably both. In particular, security related issues and software bugs need to be addressed immediately. The platform should be built upon well-established software stacks.
- Scalability. As a central repository of the NLB digital content, we envisaged the CMS to be able to scale to manage tens of millions of digital objects.
- CMIS compliance. The Content Management Interoperability Services (CMIS<sup>7</sup>) by the Organisation for the Advancement of Structured Information Standards (OASIS) provides a standard interface layer on top of existing content management systems and their existing programmatic interfaces. Compliance to CMIS would provide NLB options in future when re-considering its CMS platform.
- Full-text search. To improve on the discovery of the rich media content in the CMS, searches need to go beyond the meta-data associated with the digital objects. Moreover, the process of describing the digital objects can sometimes be labour intensive, and may not keep pace with the creation of the digital objects. Integrated full-text search capability will therefore extend the discoverability of the content.

---

<sup>7</sup> Content Management Interoperability Services (CMIS) Standard by Organisation for the Advancement of Structured Information Standards (<http://docs.oasis-open.org/cmisis/CMIS/v1.0/os/cmisis-spec-v1.0.html>)

## 4 ALFRESCO OPEN SOURCE SOFTWARE

NLB selected the Alfresco Community Edition as its content-as-a-service platform in 2010. Alfresco is an open source Enterprise Content Management (ECM) System that manages all content in an organization, and provides services and controls to manage the content.

Founded in 2005 by ECM veterans (from Documentum, Interwoven, Vignette, FileNet, etc.), the Alfresco software was designed and developed from ground-up to leverage on the modern open and established software stacks (e.g., Java, Lucene, etc.). It has been licensed under the GNU Lesser General Public License (LGPL<sup>8</sup>).

The Alfresco software is available in both the Community Edition and the Enterprise Edition. While the two editions share the same code-base, the Enterprise Edition provides commercial supports with a subscription fee, a common and established business model for open source software (e.g., Red Hat Inc for the Linux operating system, Oracle Corporation for MySQL database management system). The availability of the commercially supported option provides NLB the assurance of a smooth transition should the need arises in the future.

Extensive evaluation and benchmarking have been conducted, in line with the NLB's policy on the use of open source software (OSS) and the NLB Technical Architecture Standards. NLB regards the use of OSS in the same vein as commercial software. The use of OSS components is more ubiquitous in commercial solutions that one might have thought. Just consider how often your vendors have provided solution options based on Linux, Java, JBoss, MySQL, Apache, Axis, Lucene and Solr.

While Alfresco has published a benchmark test that exceeds 100 million objects, NLB conducted our own benchmarking exercise based on our own digital objects, set up and configuration. The benchmark was successfully completed for 5 million NLB objects in a virtual machine. We are therefore confident that it would be able to scale much more to meet NLB's future needs.

Flexible data model, extensive APIs, full-text search, high performance bulk ingest, CMIS compliance, robustness, supportability and scalability were some of the key features of the Alfresco software that were core to the content-as-a-service architecture required by NLB.

The configuration of the production CMS platform is shown in Table 1.

---

<sup>8</sup> GNU Lesser General Public License (<http://www.gnu.org/licenses/lgpl.html>)

Servers	Description
<b>Alfresco (Authoring) Server</b>	Used for bulk ingest.  Redhat Enterprise Linux Operating System Tomcat Application Server Alfresco Application
<b>Alfresco (Read only) Server Cluster</b>	Used for online service access.  Redhat Enterprise Linux Operating System Tomcat Application Server Alfresco Application
<b>DB Server Cluster</b>	Redhat Enterprise Linux Operating System MySQL DB Server

Table 1: Production CMS configuration

The Alfresco software has progressed much since 2010. The current version for the Community Edition as of this writing is 4.2.c. A cloud-based option has been added, and Alfresco Mobile iOS and Android applications launched to provide access from smart devices. Alfresco was named a ‘visionary’ in Gartner’s Magic Quadrant for Enterprise Content Management in 2012<sup>9</sup>.

## 5 SUCCESSFUL DEPLOYMENT

The platform has been successfully used in many content-rich NLB services, including Infopedia, BookSG, PictureSG, MusicSG, the Standards Portal and the Singapore Memory Portal. Table 2 provides a summary of some of these NLB services.

---

<sup>9</sup> Gartner Magic Quadrant for Enterprise Content Management 2012 (<http://www.alfresco.com/gartner>)

NLB service	Brief description	Content in CMS available for public access (as at 31 Mar 2013)
<b>Infopedia</b>	An online encyclopedia on Singapore's history, culture, people and events ( <a href="http://infopedia.nl.sg">http://infopedia.nl.sg</a> )	Infopedia articles: - Format: HTML text - #: 1,720
<b>BookSG</b>	An online collection of digitised books related to Singapore and Southeast Asia ( <a href="http://sgebooks.nl.sg">http://sgebooks.nl.sg</a> )	Digitised books: - Format: PDF - #: 13,852
<b>PictureSG</b>	An online collection of pictures, be it photographs or artworks, which provide information about Singapore ( <a href="http://pictures.nl.sg">http://pictures.nl.sg</a> )	Pictures: - Format: JPEG - #: 16,611
<b>MusicSG</b>	A non-profit digital archive set up to digitise, archive and provide access to all forms of published Singapore musical works ( <a href="http://music.nl.sg">http://music.nl.sg</a> )	Tracks: - Format: WMA - #: 4,722 Scores: - Format: PDF - #: 1,561 Lyrics: - Format: PDF - #: 2,502
<b>Singapore Memory Portal</b>	An online portal to support the Singapore Memory Project, a national initiative to collect, preserve and provide access to Singapore's knowledge materials, so as to tell a Singapore story ( <a href="http://www.singaporememory.sg">http://www.singaporememory.sg</a> )	Personal memories: - Format: various - #: 174,200 Institution memories: - Format: various - #: 43,794

Table 2: Summary of the content of the NLB services managed by CMS (not exhaustive)

## 6 BENEFITS REALISED

NLB has taken key decisions to architect its CMS as a content-as-a-service platform, and has adopted Alfresco, an open source software for the platform.

### Content-as-a-service approach

The loosely coupled, component-based architecture provided NLB the agility that enables service innovation to be built upon established, tested and repeatable CMS platform and processes.

Beyond the Responsive Web Design framework mentioned in section 2 above, the online services are able to provide the optimal user experience based on the nature of the collection.



The BookSG service integrates with the FlipViewer software that streams the PDF content with an intuitive user experience (see Figure 3).

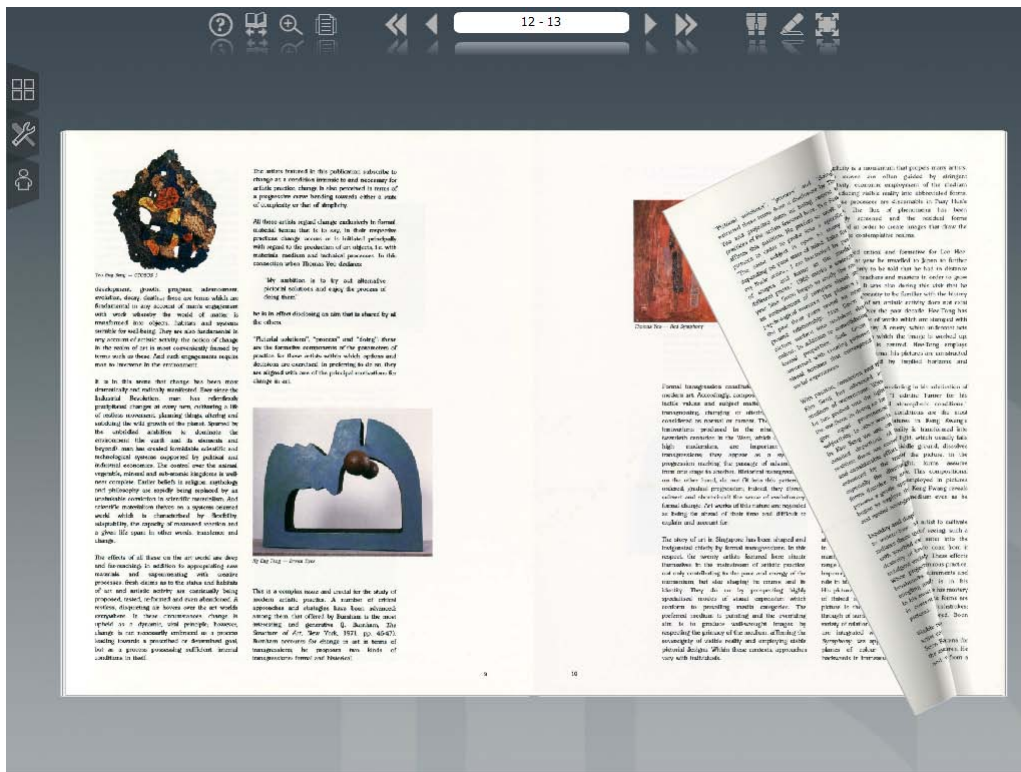


Figure 3: BookSG with its intuitive viewing experience

If the MusicSG service detects that an access is coming from mobile devices, it will provide a mobile-friendly user interface (see Figure 4), and deliver the music track from the NLB streaming infrastructure (another of the Service Enablement Layer components) that is capable of streaming audio and video to all devices.



Figure 4: MusicSG mobile-friendly interface and streaming support for all devices

The authentication to the online services is also not restricted to the out-of-the-box authentication provided in standard content management software. A good example is the Singapore Memory Portal. In order to make the contribution of personal memories as simple as possible, contributors can log in using their social account (Facebook, Google, Yahoo and Microsoft Live) on top of their NLB membership account (see Figure 5).

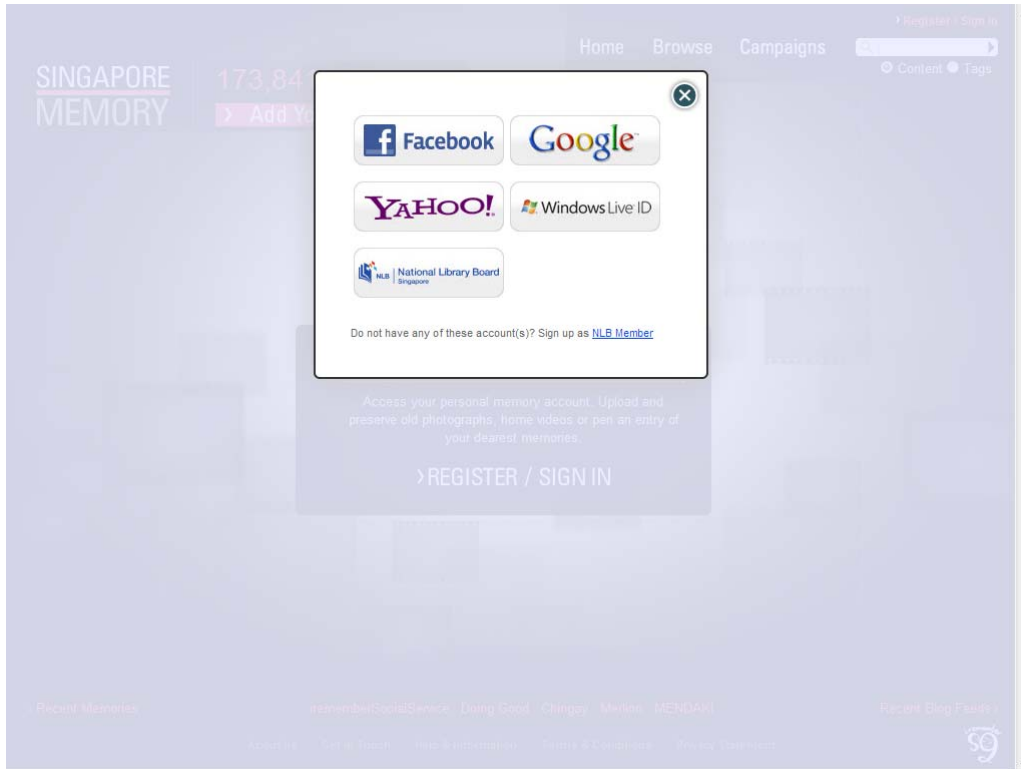


Figure 5: NLB Social Connects in the Singapore Memory Portal

While the content-as-a-service approach provides a standard set of APIs for the online services, it also streamlines the backend processes (e.g., modelling, bulk ingestion, meta-data tagging, integration with the NLB streaming infrastructure, etc.).

This architecture will also enable NLB to move to a new CMS platform more easily if the need arises in the future.

### Open source software platform

The awareness and actual use of open source software (OSS) in libraries have been increasing. Two key factors contributed to this trend:

- Maturity of the established OSS. Not all OSS are the same. In fact, only a small proportion of OSS has reached the level of maturity for production deployment in libraries. Nonetheless, such successful deployments have generated much interest, and demonstrated that they can be a good fit for certain libraries.

Examples of well-established OSS in use in libraries include Koha and Evergreen for integrated library system, VuFind for discovery, and Fedora, DSpace and Greenstone for digital repository.

- Commercial support. Due to the growing use of OSS, commercial supports can now be readily acquired. Such commercial supports are particularly important for libraries that may not have the internal resources to manage the OSS deployment and on-going operations.

The Alfresco Community Edition that NLB adopted is licensed as an OSS. We were therefore able to experiment with various configurations on virtual and physical servers, on load balancing and high availability (HA), and on ingestion and indexing, without being restraint by license cost consideration. This has allowed NLB to determine the most appropriate set up based on our current requirements, and also allow us to scale readily in the future when the needs arise.

Another benefit of the use of OSS is the community support. Online forum provides a great channel to get help on issues faced. Community contributed add-ons are also available. In particular, the bulk ingestion tool we are using is one of these community contributed add-ons.

While we do not make code changes as a matter of policy even though we are free to do so with Alfresco, the availability of the source codes, and the ability for us to scan through the codes (which we did numerous times) has allowed us to obtain a better understanding of the software, and helped us to troubleshoot and resolve issues we faced.

The Alfresco software is an extensive one with multitude of functionalities, including collaboration, record management, web content publishing management and mobile content management built on top of a common content repository. With the content-as-a-service approach, NLB uses mainly the core content repository. As the core content repository remains rather stable, we were able to upgrade our deployment to the latest version very smoothly.

## **7 LESSONS LEARNT**

The use of OSS is not a walk in the park, particularly when it is used for a critical component of the enterprise infrastructure.

### **Learning curve**

As a full-featured modern Enterprise Content Management System, Alfresco is a sophisticated software. It therefore took considerable effort for us to gain a good level of understanding of the software. Although we could have work with an experience vendor to quicken the first implementation, on hindsight, the deep knowledge we gained would stand us in good stead in the long run.

We conducted a health-check exercise with an external consultancy in December 2011. The exercise was very fruitful as NLB by then has accumulated sufficient knowledge in Alfresco and a set of ‘difficult’ issues to work with the consultants.

## **Mindset change**

Another significant change is the result of the decision to architect the CMS as a content-as-a-service re-usable component. Previously each project will provide for its own CMS. The focus now instead is in integrating with the central CMS platform.

The initial sense of the ‘lost of control’ from the project managers and the IT architects of these projects was apparent. Nonetheless, as the CMS platform matures, and the process of integration better understood, the content-as-a-service approach yielded many benefits (some of which described in Section 6 above). From an organisational perspective, once we managed to implement a new capability through the CMS, we can immediately apply it to all future projects with similar requirements. Examples include the on-the-fly watermarking of images and PDF, the handling of limited preview of PDF documents, and many more.

## **Resources**

The adoption of an OSS for a key business requirement demands the corresponding commitment of resources. While the community support is there, it is inherently different from that of commercial support with stipulated service level agreements (SLAs). Building the internal competency and capability to support the OSS was critical in providing the necessary assurance to the various stakeholders.

## **No Code Changes**

The out-of-the-box user interface for the maintenance of the meta-data associated with the digital objects turned out to be less streamlined than the existing meta-data tagging system (GT System). In line with our policy of not making code changes to our OSS deployment, we took the opportunity to revamp the GT System to integrate with the CMS (GT.Net System).

The discipline to not make code changes has enabled NLB to continue to upgrade and up-keep our version of the Alfresco software with minimal efforts.

## **Building our own API layer**

To further improve on portability, we have also developed a number of our own APIs (in the form of web services) on top of the APIs provided by Alfresco. Examples of these NLB web services include search, retrieve metadata, retrieve content, etc.

## **8 CONCLUSION**

NLB has selected the Alfresco Community Edition, an OSS for its content-as-a-service platform. It is a critical component of the overall IT architecture and application infrastructure, and has been successfully implemented to enable many innovative and popular content-rich NLB services.